INDIAN SPACE RESEARCH ORGANISATION

IPRC-MAHENDRAGIRI



**INPLANT TRAINING REPORT**

(26.06.2023 TO 07.07.2023)

**Prepared by**

**NABOTH DEMETRIUS R**

**MOHAMED ASLAM K**

**INPLANT TRAINEE (Data Science)**

Guided By: Approved By:

**Anitha T,**

**Mahesh Kumar S**

**COWAA, MSA GH, MSA**

COMPUTERISATION WORKING OF ADMINISTRATIVE AREA

MANAGEMENT SYSTEM AREA

ISRO PROPULSION RESEARCH COMPLEX

MAHENDRAGIRI -627 133

JULY-2023

Table of Contents

[1. Introduction 4](#_Toc139553951)

[2. Database Management Systems 4](#_Toc139553952)

[2.1. RDBMS: 5](#_Toc139553953)

[2.2. ACID: 5](#_Toc139553954)

[2.3. RAID: 6](#_Toc139553955)

[2.4. OLAP: 6](#_Toc139553956)

[2.5. Processing of a transaction: 7](#_Toc139553957)

[2.6 Possible Applications: 7](#_Toc139553958)

[3. Firewall 7](#_Toc139553959)

[3.1. Web Filtering 8](#_Toc139553960)

[3.2. Centralized Management 8](#_Toc139553961)

[3.3. MAC Binding 8](#_Toc139553962)

[3.4. Privacy and Backup 8](#_Toc139553963)

[3.5. Possible Application 9](#_Toc139553964)

[4. Network 9](#_Toc139553965)

[4.1. Logical Grouping 9](#_Toc139553966)

[4.1.1. Server switches: 9](#_Toc139553967)

[4.1.2. Employee: 9](#_Toc139553968)

[4.1.3. Audio: 10](#_Toc139553969)

[4.1.4. Video: 10](#_Toc139553970)

[4.2. Procurement 10](#_Toc139553971)

[4.3. Possible Applications 11](#_Toc139553972)

[5. Telephone Exchange 11](#_Toc139553973)

[5.1 Components 11](#_Toc139553974)

[5.1.1. EPABX 11](#_Toc139553975)

[5.1.2. MDF 11](#_Toc139553976)

[5.1.3. Jumper Cables 11](#_Toc139553977)

[5.1.4. Armored Jelly Fish Cables 12](#_Toc139553978)

[5.1.5. Junction Box 12](#_Toc139553979)

[5.1.6. Landline 12](#_Toc139553980)

[5.2. FWT 12](#_Toc139553981)

[6. MET visit 12](#_Toc139553982)

[6.1 Engine 13](#_Toc139553983)

[6.2 PLC 13](#_Toc139553984)

[6.2.1 Abort 13](#_Toc139553985)

[6.2.2 Interlocks 14](#_Toc139553986)

[6.2.3 Emergency 14](#_Toc139553987)

[6.2.4 Implementation 14](#_Toc139553988)

[6.3 Operating System: 14](#_Toc139553989)

[7. Dimensionality Reduction 14](#_Toc139553990)

[7.1 Principal Component Analysis 15](#_Toc139553991)

[7.2 Linear Discriminant Analysis 15](#_Toc139553992)

[7.2.1 PCA VS LDA 16](#_Toc139553993)

[7.3 Factor Analysis 16](#_Toc139553994)

[7.4 Reducing dimensions using Neural Network 17](#_Toc139553995)

[7.4.1 Auto encoders 17](#_Toc139553996)

[7.4.2 Encoder-Decoder Networks 17](#_Toc139553997)

[7.4.3 Dimension Reducing Layers 17](#_Toc139553998)

[8. Tasks Completed 18](#_Toc139553999)

[8.1 TASK 1 - Connectivity log analysis: 18](#_Toc139554000)

[8.2 TASK 2 - Principal Component Analysis: 20](#_Toc139554001)

[8.3 TASK 3 - Linear Discriminant Analysis: 22](#_Toc139554002)

[8.4 TASK 4 – Training Neural Networks in Tensor-flow Playground: 25](#_Toc139554003)

[9. Conclusion: 28](#_Toc139554004)

# 1. Introduction

This report provides an overview of our in-plant training experience at the ISRO Propulsion Complex in the field of data science. The training program, which took place at IPRC's facility in Mahendragiri, aimed to enhance my understanding of the need for using data science techniques and technologies in analyzing the data from various test sources. Also, we have learned various technologies and their applications that will be a pioneering knowledge to understand different kinds of data to get creative ideas on how to implement various Machine Learning Techniques in different kinds of applications.

ISRO is renowned for its significant contributions to India's space program and has been at the forefront of scientific advancements. The organization utilizes various software and hardware technologies to gather, analyze, and interpret vast amounts of data to drive critical decisions and facilitate groundbreaking discoveries. IPRC stands for the Indian Space Research Organization Propulsion Complex. It is an organization within the Indian Space Research Organization (ISRO) that specializes in the development and production of propulsion systems for space missions. IPRC is responsible for designing, testing, and manufacturing liquid propulsion engines, stages, and control systems for various launch vehicles, satellites, and spacecraft. It plays a crucial role in advancing India's space exploration capabilities by ensuring reliable and efficient propulsion systems for ISRO's missions.

# 2. Database Management Systems

Database systems refer to software applications or systems that are designed to efficiently store, manage, and retrieve large amounts of structured data. As the Facility needs to store a massive amount of data for each and every test conducted, it is necessary to store the data efficiently in order to be able to analyze it when needed. In IPRC a well-established Relative Database Management System is used to store the test data. Not only test data but also it is needed to store information about the payments made either internally through Canteens, Stores, etc., or externally through vendors from outside the complex. Maintaining these data is very crucial as each payment made should be sent to NIC in order to validate and sanction the amount needed for the payment. Now let’s further look into how the Database is maintained for this large organization. Various concepts like ACID, RAID, etc. are used to achieve this.

## 2.1 RDBMS:

A relational database system is a type of database management system (DBMS) that organizes and stores data in a structured manner using tables, which are composed of rows and columns. There are 13 rules through which we can construct an RDBMS.

* Information Rule
* Guaranteed Access Rule
* Systematic Treatment of Null Values
* Active Online Catalog Based on the Relational Model
* Comprehensive Data Sublanguage Rule
* View Updating Rule
* High-Level Insert, Update, and Delete Rule
* Physical Data Independence
* Logical Data Independence
* Integrity Independence
* Distribution Independence
* Non Subversion Rule
* Dynamic Online Catalog Based on the Relational Model

These 13 rules provide us with the design and framework to implement an RDBMS.

**Used in IPRC:** Canteen Management System, Salary Management System, etc. **Note:** More than 2400 tables are used in IPRC’s database management system.

## 2.2 ACID:

ACID is an acronym that stands for Atomicity, Consistency, Isolation, and Durability. These properties are essential for ensuring the reliability, integrity, and consistency of transactions in a database management system (DBMS).

* Atomicity
* Consistency
* Integrity
* Durability

It is highly important in IPRC to adapt the ACID properties in order to provide a well-working clean database management system that manages data efficiently.

## 2.3 RAID:

For a large organization, it is necessary to have multiple data storage disks merged into a single one so that it can be treated as a single storage unit at the time of retrieval and storing the data. RAID stands for Redundant Array of Independent Disks. It is a data storage technology that combines multiple physical hard drives or solid-state drives (SSDs) into a single logical unit for improved performance, fault tolerance, or both. RAID offers various levels or configurations, each with its own advantages and trade-offs. In IPRC RAID-6 is used in systems that have the following features,

* RAID 0: Data striping across multiple drives for improved performance, but no data redundancy.
* RAID 1: Mirroring data on two drives for redundancy and increased data reliability.
* RAID 2: Striping at the bit level with dedicated error correction, rarely used in practice.
* RAID 3: Byte-level striping with dedicated parity disk for improved performance and fault tolerance.
* RAID 4: Block-level striping with a dedicated parity disk for enhanced performance and fault tolerance.
* RAID 5: Block-level striping with distributed parity for both performance and fault tolerance.
* RAID 6: Block-level striping with dual distributed parity for higher fault tolerance than RAID 5.

## 2.4 OLAP:

OLAP stands for Online Analytical Processing. OLAP (Online Analytical Processing) focuses on a complex, multidimensional analysis of data for decision support, while OLTP (Online Transaction Processing) focuses on real-time transactional processing and management of day-to-day operational data. Though we have a large variety of data, we store them in a single place in the main building of IPRC. This place where we store all the data in a single place is known as a **Data warehouse**. This server also has a redundant server in the IT services building.

## 2.5 Processing of a transaction:

For example, when we have to record a medical payment bill, we have to send it to the NIC server in order to get sanctioned for the amount spent. So the following steps can be followed,

* We get the bill details from the user through the web application.
* We update the details on our IPWS (ISRO PFMS Web Server).
* Then we validate the details provided by the mentioned medical personnel.
* Then we convert the details into the form of an XML file.
* Then we encrypt the XML as the XML is going to be sent through the public server to the PFMS (Public Financial Management System).
* Then the XML is decrypted in the NIC server and the needful is done.

## 2.6 Possible Applications:

This massive stored data in the systems can be used to train efficient models based on the need required. In the case of Engine test datasets, we can reduce the dimensions as there are a lot of parameters being recorded in it which will be not much efficient to train models.

# 3. Firewall

A firewall is crucial for IPRC to protect its network infrastructure and sensitive data from unauthorized access and potential cyber threats. As the data stored is very crucial and must not be let down into the hands of potential threat-causing agents, it is needed to possess a well-established Firewall. In IPRC Sophos firewall is used to protect networks and devices from various threats, including malware, unauthorized access, and data breaches.

## 

## 3.1 Web Filtering

It includes web filtering capabilities to control and monitor internet access. It can block access to malicious or inappropriate websites, ensuring a safer browsing experience for users. In IPRC we restrict access to websites that may be a threat to the data. The harmful websites will be found by the National Investigation Agency. These websites are harmful and may be a major threat to confidential data. These websites will be added on the web filtering tab on Sophos so they won't be accessed by the user.

## 3.2 Centralized Management

Sophos Central, a cloud-based management platform, allows administrators to centrally manage and monitor multiple Sophos Firewalls from a single console. It simplifies deployment, configuration, and ongoing management of the network security infrastructure. We can also limit the bandwidth based on each user. In IPRC all the internet facilities are centralized through Sophos. Every user has their credential to access the internet. In the development and management of data, all personal computers are isolated from the internet and they are connected through an intranet. This isolation will protect confidentiality and information and reduce the threat.

## 3.3 MAC Binding

We can use this technique to bind the Mac address of a system to a particular firewall profile so that the user cannot access the network through any other system. This could be helpful in avoiding insecure systems to be connected to the network. Also, we can have a White-list in which we specify the allowed MAC IDs so that no new system will be able to access the network which cripples the ability to hack or harm our Network.

## 3.4 Privacy and Backup

Every search made or every action made by the network must be tracked and stored in a Database. This is necessary so that during times of any attacks we would be able to track the attack through the history we have stored through the backups made. Also, logs of 2 years must be stored in the database as per IT Security Norms.

## 3.5 Possible Application

We can train a model that is trained on the recorded logs by this we can be able to detect anomalies in the network and tag them as suspicious. These suspicious connections can be scanned for any possible threats and actions can be taken accordingly.

# 4. Network

Networking refers to the practice of connecting computers, devices, or systems together to facilitate communication and data exchange. It enables the sharing of resources, such as files, printers, and internet connections, among multiple users or devices.

## 4.1 Logical Grouping

Logical grouping in networking refers to the practice of organizing devices, resources, or users into logical units or segments based on certain criteria or requirements. It helps in simplifying network management, improving security, and optimizing network performance. These logical groupings are done through Virtual LANs (VLANS). In IPRC VLANs are divided into four that is

* Server Switches
* Employee
* Audio
* Video

## 4.1.1 Server switches:

Server switches in VLANs are switches specifically designed to connect servers within a network and to facilitate their communication within specific VLANs. VLANs (Virtual Local Area Networks) allow for the logical segmentation of a network, enabling different groups of devices to be grouped together virtually, regardless of their physical location.

## 4.1.2 Employee:

An employee VLAN is a specific VLAN (Virtual Local Area Network) created to group together the network ports or devices used by employees within an organization. It provides a logical segmentation of the network, allowing for better management, security, and control of employee network traffic. In IPRC all the employees are connected through this network and access the database of the IPRC.

## 4.1.3 Audio:

An audio VLAN is a specific VLAN (Virtual Local Area Network) created to handle audio traffic within a network. It is designed to prioritize and segregate audio data to ensure optimal performance, quality, and reliability for audio applications. In IPRC it is used for audio streaming, video conferencing, or other audio-intensive applications.

## 4.1.4 Video:

A video VLAN is a specific VLAN (Virtual Local Area Network) created to handle video traffic within a network. It is designed to prioritize and segregate video data to ensure optimal performance, quality, and reliability for video applications. In IPRC it is used for video conferencing, video streaming, and surveillance systems.

## 4.2 Procurement

Two important terms are to be noted that is **Indent** and **Tender**. Indent is the need raised by the engineers when they see there is a need for any equipment. This Indent is sent to the purchase committee where the committee checks if the need is valid or not. The engineer who raises the indent has to select two purchase selectors and two administrative officers. The approval of the selected officers is needed to further move the indent into the market. Tender is when the Indent is sent into the market for the parties to bid. The party which bids the lowest bid is selected. Then if the bid is higher than the estimated amount it is again sent to the engineer to decide if they can proceed or cancel the bid. This is how equipment’s or any appliances are bought in IPRC. So a well-established seamless network is established for raising indents and tenders. Also, it is made easier for the authorities to validate the indents through interactive dashboards.

## 

## 4.3 Possible Applications

On looking at these processes we could speculate the possibility that it will be possible for us to create **a model that is trained on valid and invalid indents** based on various parameters which will be helpful for the purchase committee to accept or reject an Indent.

# 

# 5. Telephone Exchange

Telephone exchange includes the intercom facility for the engineers and employees to communicate with each other through a private communication service. Not only within themselves but also they are able to communicate outside.

## 5.1 Components

## 5.1.1 EPABX

EPABX stands for Electronic Private Automatic Branch Exchange. It is a telephony system used in organizations to manage incoming and outgoing telephone calls, as well as internal communication. EPABX systems provide various features such as call routing, call forwarding, voice mail, conference calling, and call transfer, enabling efficient communication within an organization. It is able to serve 24 numbers per single analog/digital card.

## 5.1.2 MDF

MDF, or Main Distribution Frame, is a central point in a telecom network where multiple incoming and outgoing communication lines are terminated, connected, and managed. It serves as a hub for connecting external lines from the service provider to internal lines within a building or facility.

## 5.1.3 Jumper Cables

Telecom jumper cables are short cables used to establish connections between different points within a telecommunication system, typically connecting equipment, modules, or ports for seamless communication and signal transmission.

## 5.1.4 Armored Jelly Fish Cables

These are similar cables but they have enhanced protection that allows them to be laid on any place. It is used to establish connections between points outside the system.

## 5.1.5 Junction Box

These boxes are used to connect the cables and then provide them to the respective endpoints where the service can be used.

## 5.1.6 Landline

A landline is a telephone that transmits signals converted from audio data through physical media, such as wire or fiber optic cable, rather than through wireless transmission

## 5.2 FWT

FWT (Fixed Wireless Terminal) phones are devices that enable wireless communication for voice calls using cellular network technology. They serve as a bridge between traditional landline telephones and mobile networks, allowing users to make and receive calls using standard landline phones but wirelessly. FWT phones typically consist of a base unit that connects to the cellular network and provides the necessary interfaces to connect landline phones for voice communication.

# 

# 6. MET visit

MET stands for Main Engine Test facility. There is a control center where the parameters of the test are controlled and the testing is done far from the control center. This is to ensure that, even if there happens to be a failure in the test, there will be no casualties.

## 6.1 Engine

The engine has three main stages that is,

* Starting stage
* Steady stage
* Shut down stage

Based on these stages the controllable parameters like the flow and pressure is controlled from the control center. There are two valves namely Electro Pneumatic Valves and Control valves to control the flow. EP valves are in either ON state or in OFF state. By using Control valves to open or close the valves in a certain percentage level. The parameters are recorded during the test using **Transmitters** and **Transducers.** For the ON state of the EP valves 24V is given. Control valves are controlled by Analogs.

## 6.2 PLC

PLC stands for Programmable Logic Control. It is controlled either digitally or by Analogs. This PLC is used to remotely control the parameters in the Test stand from the control center during the hard tests. We can manually turn on EP valves or control the flow in control valves. But the problem with doing this manually is that it is not practically possible for us to control a lot of parameters that have to change in microseconds manually. So we have a technique known as **Auto Sequencing.** The technical team works with the Mechanical team to get the knowledge about when and how the pipes should be controlled for the respective tests. By using this knowledge we form out a csv in which we numerically represent the controlling of the valves. This is fed to the PLC and the PLC uses this csv data to control the Test Facility parameters.

## 6.2.1 Abort

This mechanism acts when there is a deviation of a parameter from the original values. This is as even a slight deviation could lead to heavy damage if the plant. These are monitored by maintaining threshold values for each parameter. During abort, the current executing sequence stops and the abort sequence starts to execute. This isknown as **Auto Abort.** But we also have a **Manual Abort** in which we manually interrupt the sequence using a switch by monitoring the important factors on screen.

## 6.2.2 Interlocks

Interlocks are set in the chambers where the liquids are filled. PLC monitors the level of filling and if by accident if the level surpasses the threshold value, then the Interlocks stop the filling, thus avoiding any damage.

## 6.2.3 Emergency

We always do have redundant systems for every entity we have including storage systems, power units, etc. as all the tests are very critical and hazardous. Similarly we also have a redundant PCA to use if in case the main PCA gets disrupted. But if both of them are unable to function then the Emergency mechanism comes into action which is present in the facility. This shuts down the facility as fast as possible by stopping flow and shutting down the systems.

## 6.2.4 Implementation

The PLC is implemented by using,

* Functional Block Diagram - Graphical Representation
* Structured Test - Scripting
* Ladder Diagram - Graphical Representation (old)
* Instruction test - Scripting (old)

Basically these are in C language and the visual representations are in Python.

## 6.3 Operating System:

The OS used in the Control center is APROL which is a Linux distribution developed by B&R which is specially designed for PLC systems. Red Hat Linux is used for Data **Acquisition, Storing and Processing** of all the parameters in real time.

# 7. Dimensionality Reduction

It's a way of reducing the size of the data set by reducing the number of columns that are Features of the data. There are many methods to reduce the dimension some of the important techniques are

* Principal Component Analysis (PCA)
* Linear Discriminant Analysis (LDA)
* Factor Analysis (FA)
* Reducing dimension by Neural Networks (NN)

## 7.1 Principal Component Analysis

Principal components analysis (PCA), is a statistical technique used to analyze and simplify a dataset by transforming it into a new set of variables called principal components. Each principal component is a linear combination of the original variables in such a way that they capture the maximum amount of variance in the data.

The goal of PCA is to reduce the dimensionality of the dataset while retaining as much information as possible. By representing the data in terms of its principal components, which are orthogonal to each other, PCA helps to uncover the underlying structure and patterns in the data.

The first principal component accounts for the largest possible variance in the dataset, followed by the second principal component, and so on. Each subsequent principal component explains a decreasing amount of variance compared to the previous one. Therefore, by selecting a subset of the principal components that explain a significant portion of the variance, one can effectively reduce the dimensionality of the dataset.

Principal components are computed by finding the eigenvectors and eigenvalues of the covariance matrix or correlation matrix of the original dataset. The eigenvectors represent the directions in the feature space along which the data varies the most, while the eigenvalues indicate the amount of variance explained by each eigenvector.

PCA has various applications in data analysis, such as dimensionality reduction, data visualization, noise reduction, and feature extraction. It is widely used in fields like finance, image processing, genetics, and social sciences to uncover meaningful patterns and simplify complex datasets.

## 7.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a statistical technique used for dimensionality reduction and classification tasks. It aims to find a linear combination of features that maximizes the separation between different classes in the data. LDA is commonly used in machine learning and pattern recognition to analyze and classify datasets.

The goal of LDA is to transform the original feature space into a lower-dimensional space while preserving the discriminatory information between different classes. By projecting the data onto a set of orthogonal axes called discriminant axes, LDA seeks to minimize the within-class scatter and maximize the between-class scatter.

**Note:** Data should have class or target value to perform LDA

## 7.2.1 PCA VS LDA

* To perform Principal Component Analysis we don't require data with target values but for LDA data should have target values.
* Data is reduced through PCA is applied on unsupervised algorithms but for LDA is applied on supervised algorithms.
* In PCA dimensions are reduced in such a way that maximum variance is covered but in LDA data is separated by maximum linear separation between them.
* PCA helps to identify the important features but in LDA it improves the performance of classification algorithms.

## 7.3 Factor Analysis

Factor analysis is a statistical technique used to identify underlying latent variables, known as factors, which explain the patterns of correlation among a set of observed variables. It aims to uncover the underlying structure or dimensions in a dataset by reducing the number of variables and identifying the common factors that contribute to their variation.

The key idea behind factor analysis is that observed variables are not only influenced by specific factors of interest but also by random or error factors. The goal is to separate the systematic variance (explained by the factors) from the random variance (explained by error factors).

## 7.4 Reducing dimensions using Neural Network

Reducing dimensions using neural networks typically involves employing techniques such as auto encoders or dimensionality reduction layers. Here's a general overview of how dimensionality reduction can be achieved using neural networks.

## 7.4.1 Auto encoders

Auto encoders are neural networks designed to reconstruct their input data. They consist of an encoder and a decoder. The encoder maps the input data to a lower-dimensional representation, while the decoder reconstructs the original input from this lower-dimensional representation. By training the auto encoder to minimize the reconstruction error, the encoder effectively learns a compressed representation of the input, resulting in dimensionality reduction.

## 7.4.2 Encoder-Decoder Networks

Another approach involves using encoder-decoder networks, also known as variation auto encoders (VAEs). VAEs add a probabilistic component to auto encoders, which allows them to generate new samples similar to the input data. VAEs can be used for dimensionality reduction by training the encoder to map the input data to a lower-dimensional latent space with specific properties, such as a Gaussian distribution.

## 7.4.3 Dimension Reducing Layers

Some neural network architectures, such as the U-Net or the Inception architecture, include built-in dimensionality reduction layers. These layers reduce the spatial dimensions of the input data while preserving relevant features. They are commonly used in tasks such as image segmentation or object detection.

# 8. Tasks Completed

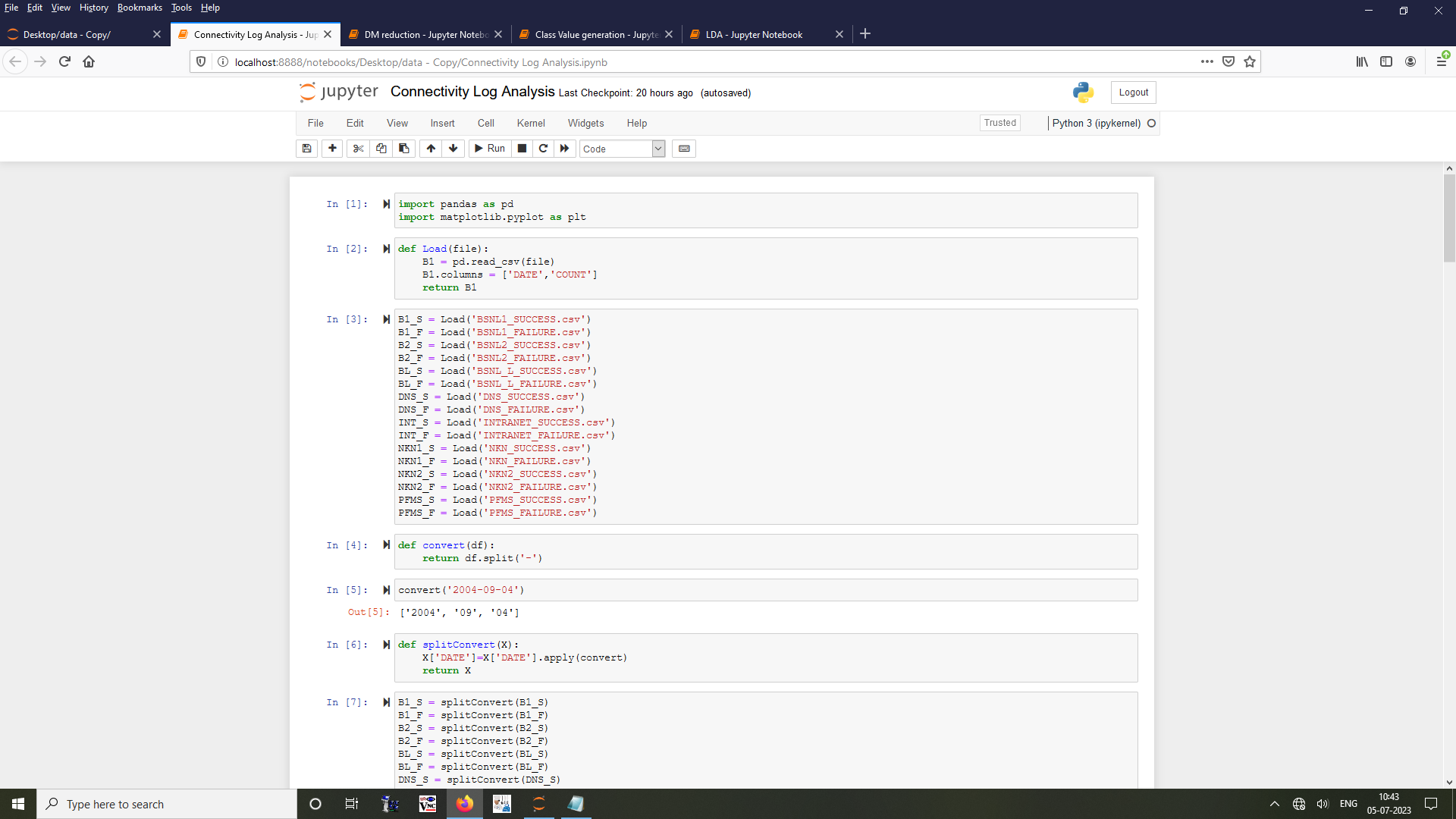
## 8.1 TASK 1 - Connectivity log analysis:

**Aim:**

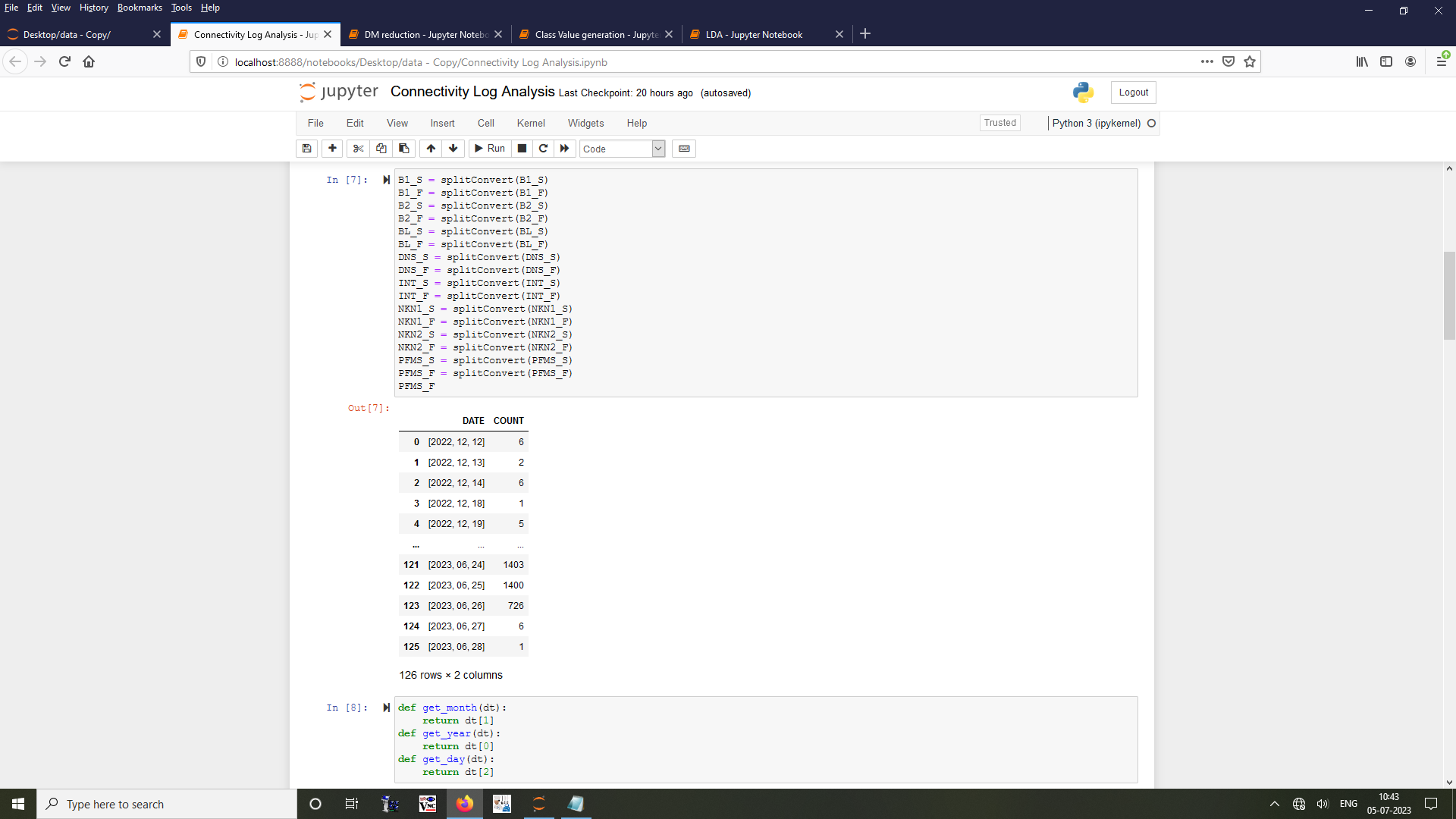
To analyze status of the **I**SRO **P**FMS **W**eb Interface **S**ystem-**P**ublic **F**inancial **M**anagement **S**ystem (IPWS-PFMS) connectivity monitoring system to address (Either success or failure) and visualize them for a given month.

**Algorithm:**

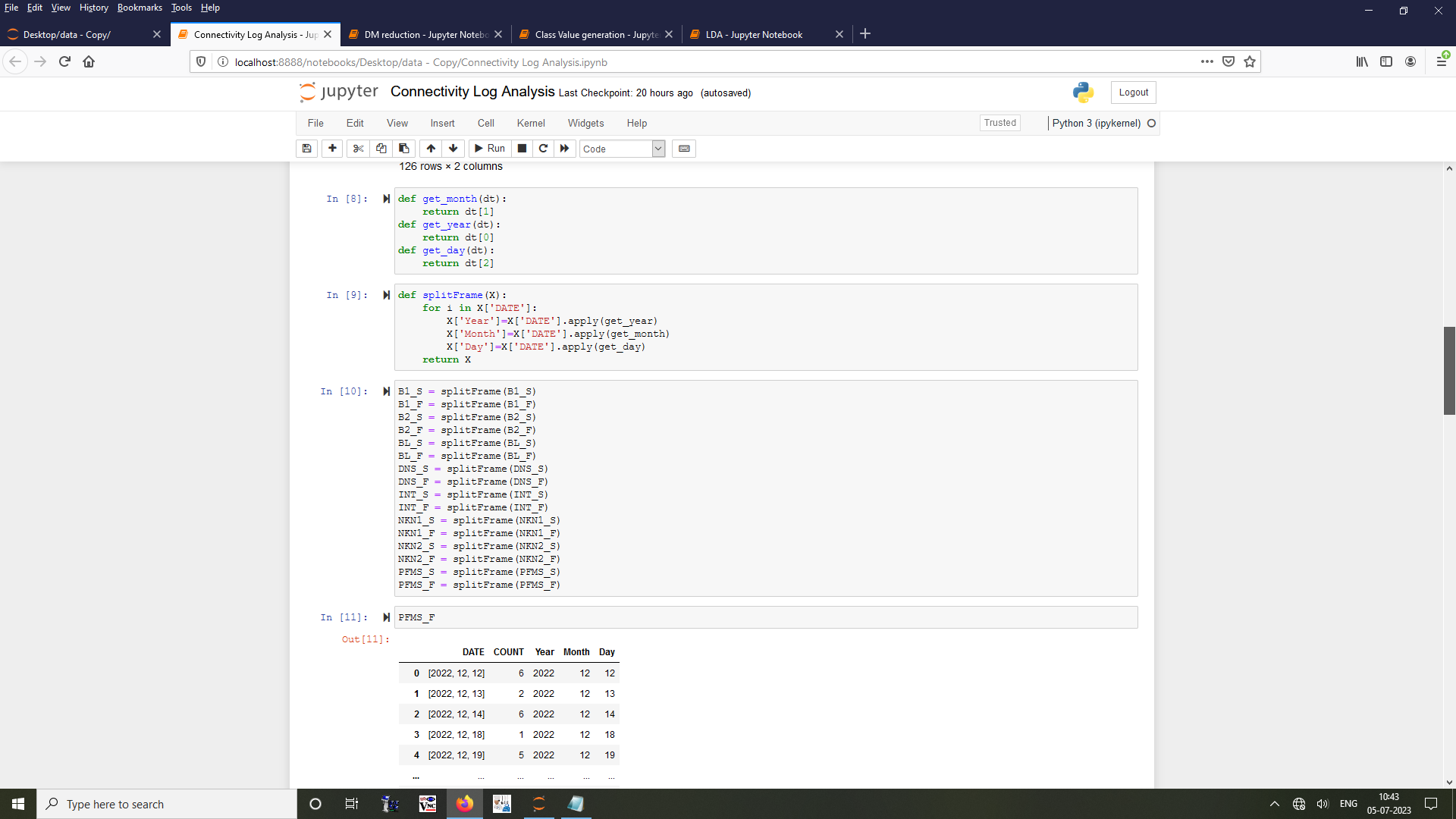
1. Load the necessary files.



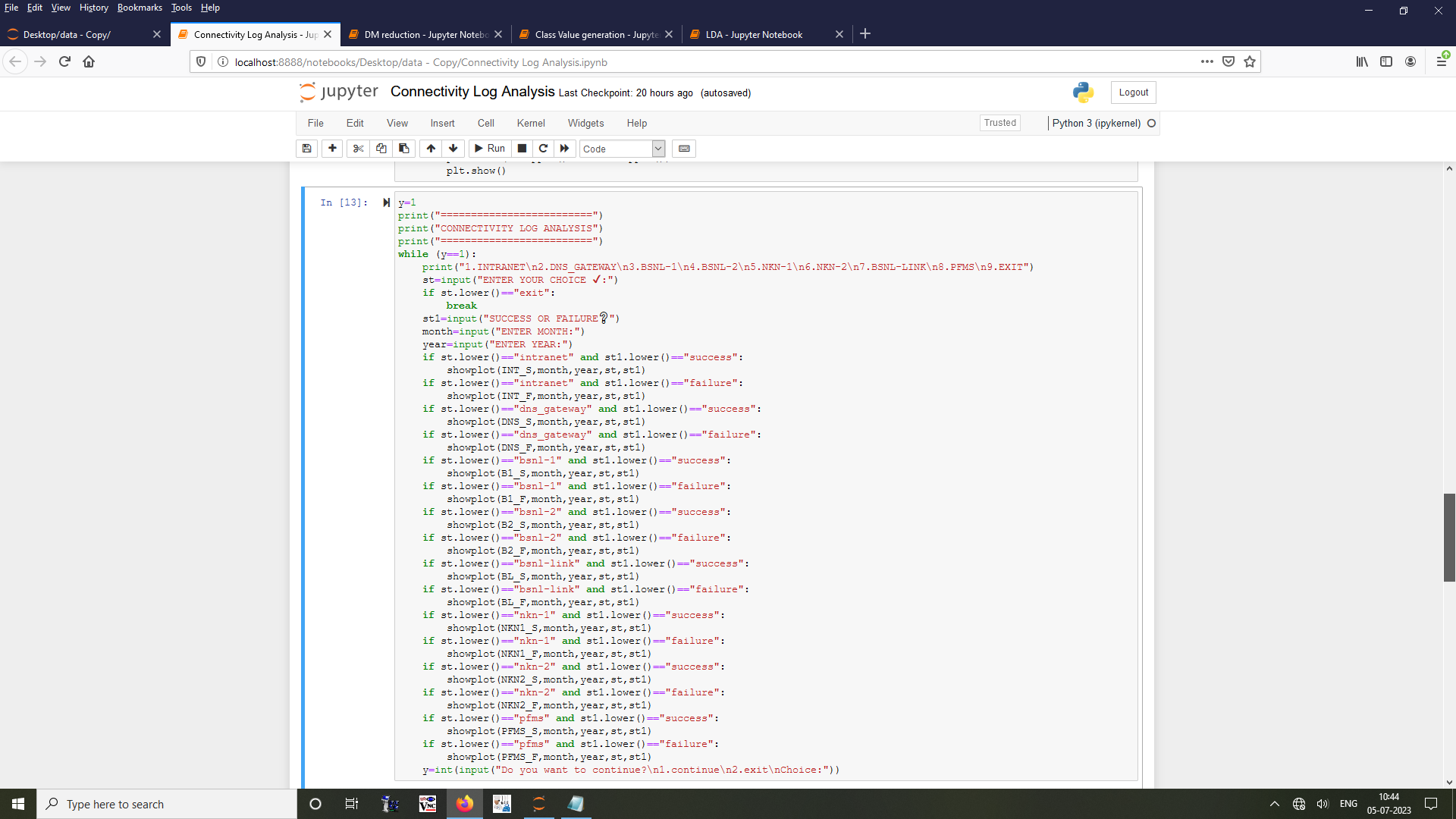
1. Split the date into day month year.



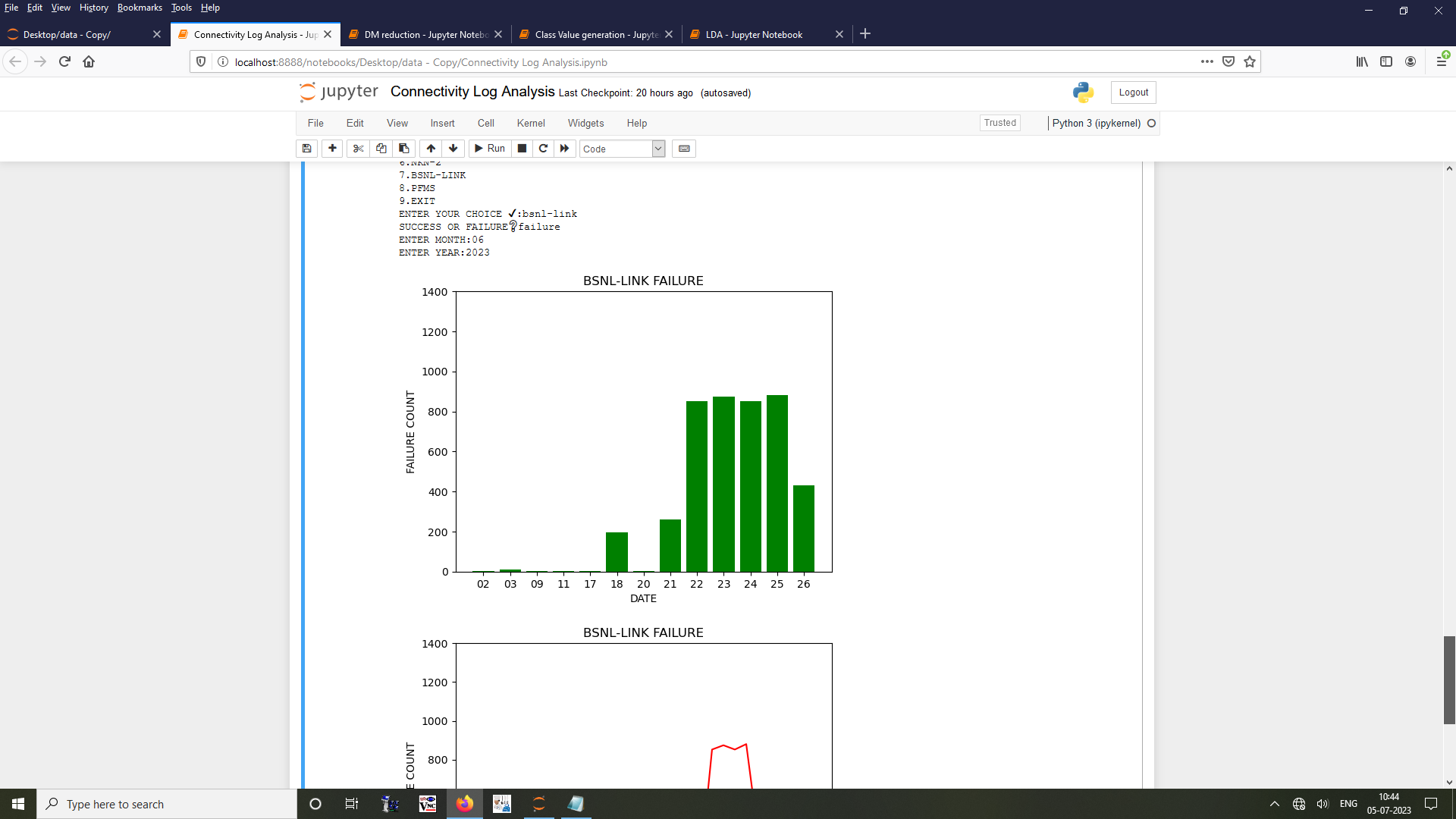
1. Assign each to its respective columns.



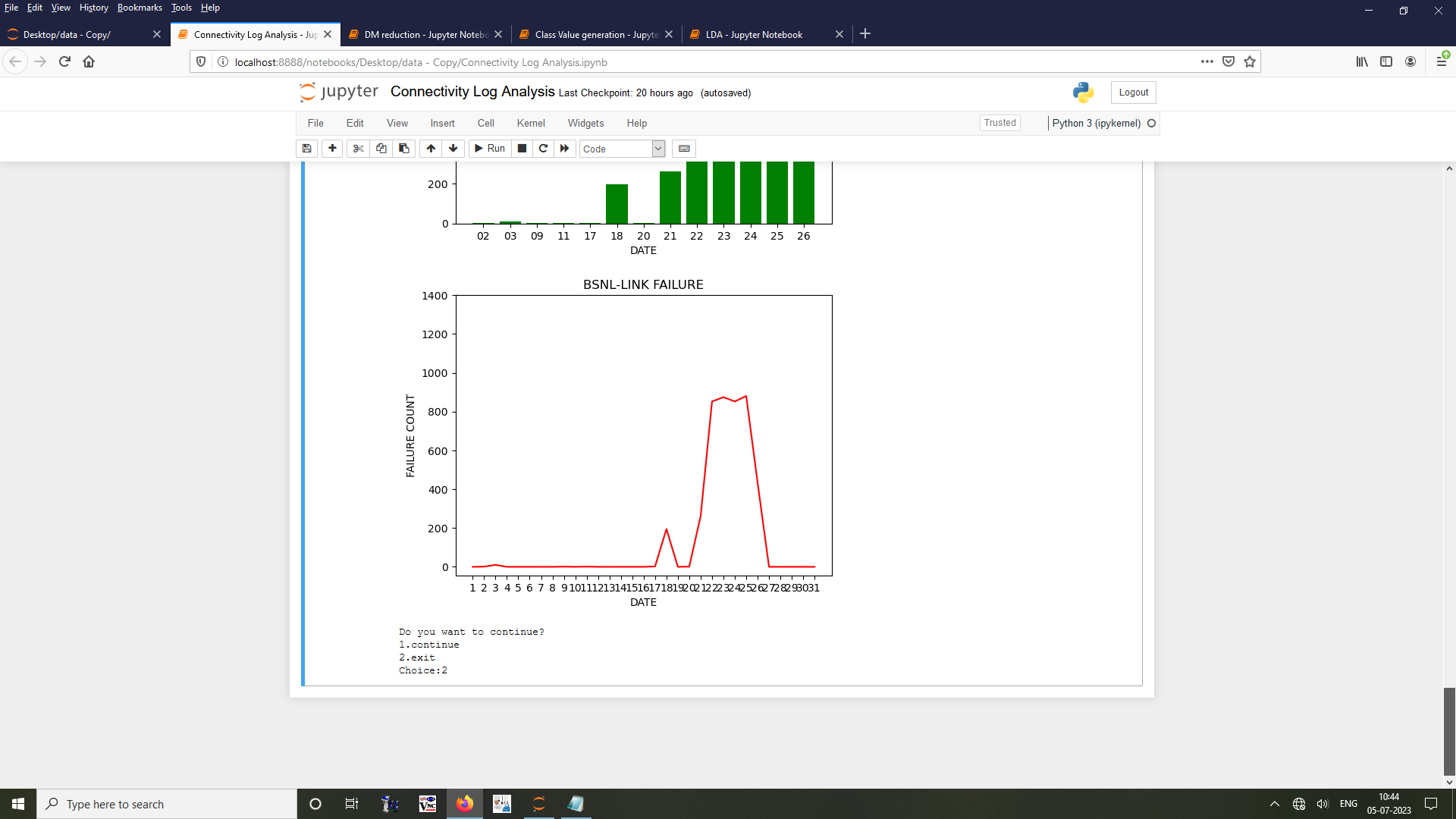
1. Get the IP, status, month, year.



1. Visualize the count of the required status using a bar chart.



1. Using a line plot visualize the fluctuations for the complete month.



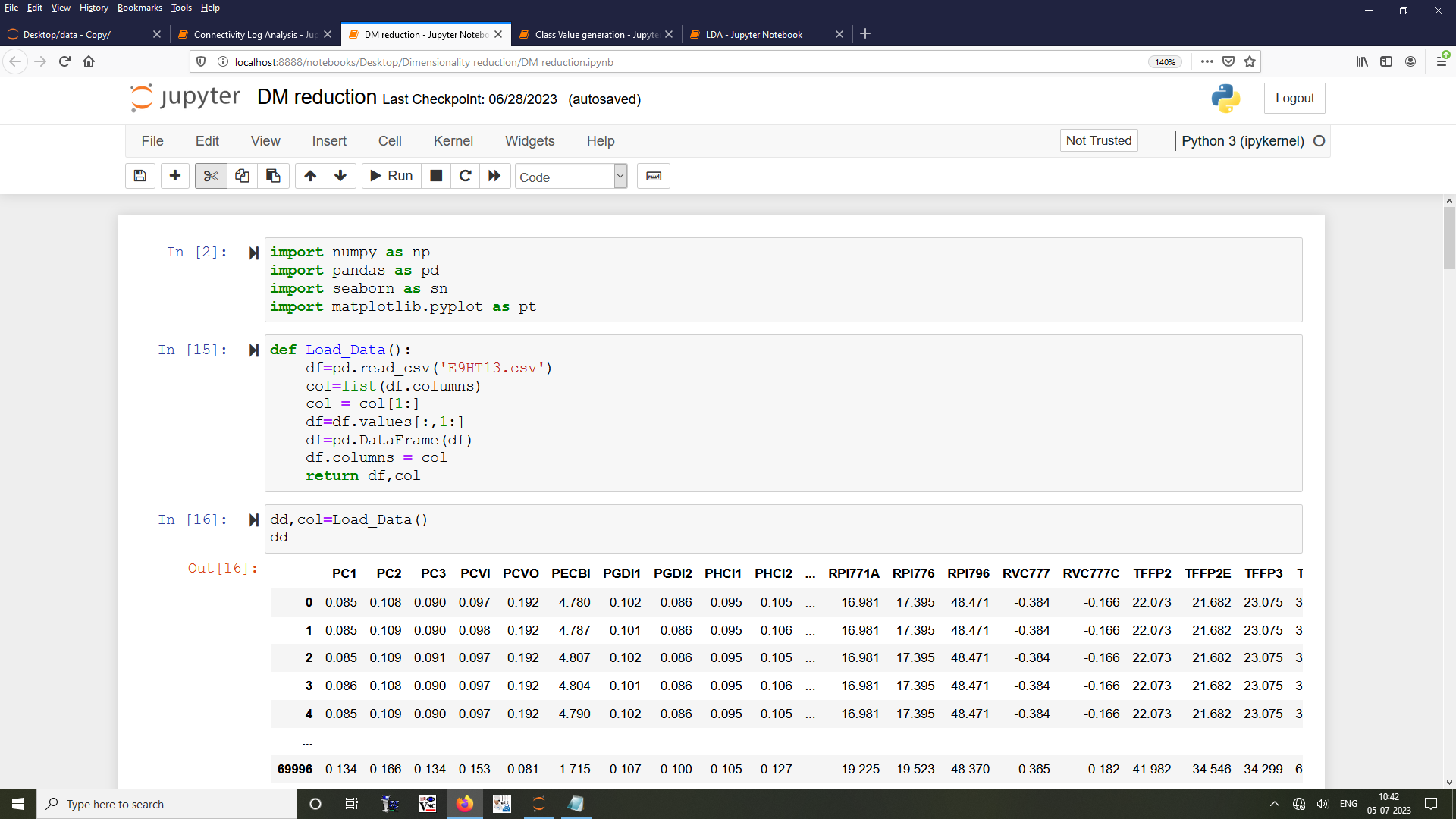
## 8.2 TASK 2 - Principal Component Analysis:

**Aim:**

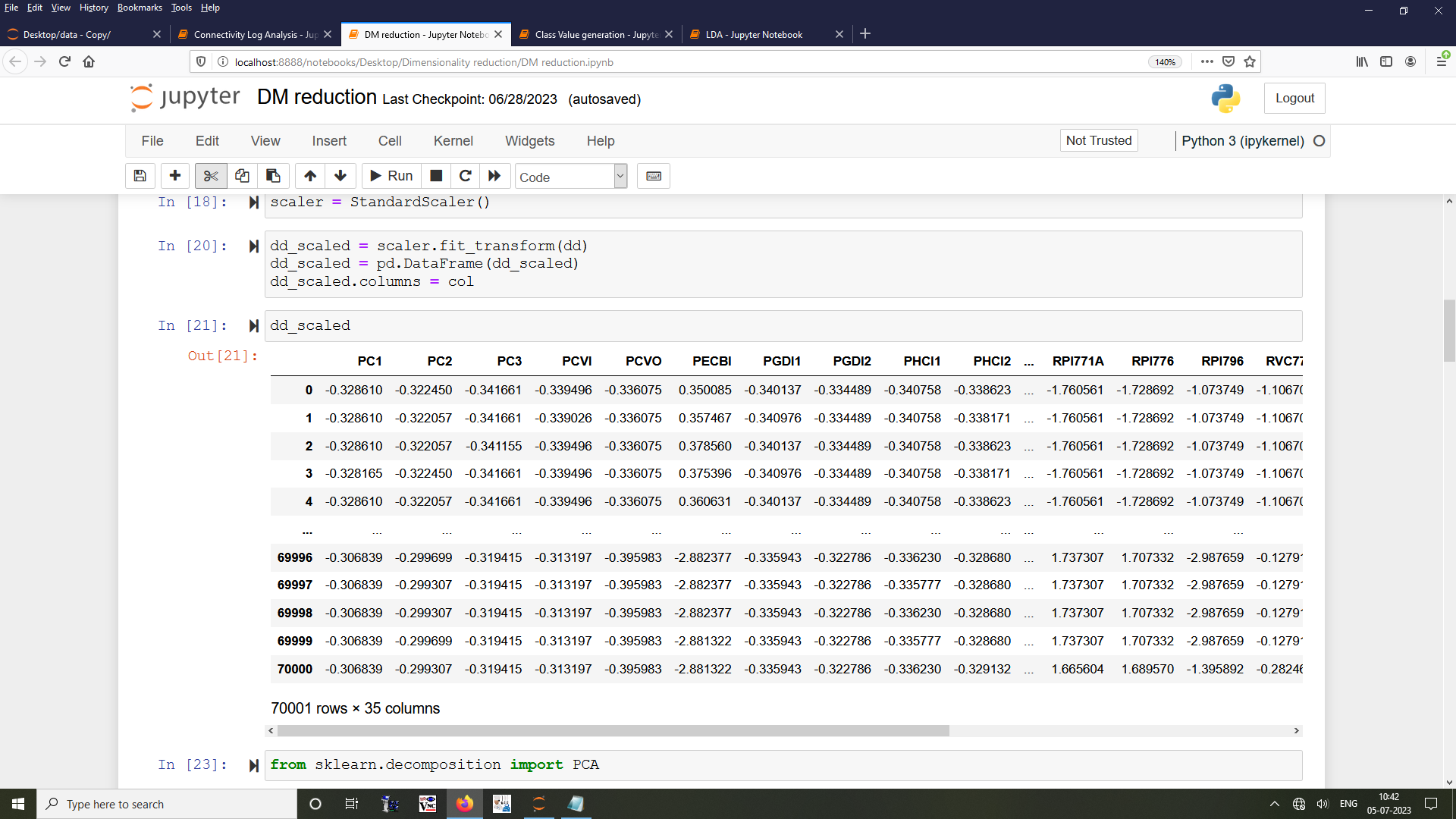
To perform dimensionality reduction using principal component analysis (PCA) on given engine data set and find the principal components which explains variance of 90%.

**Algorithm:**

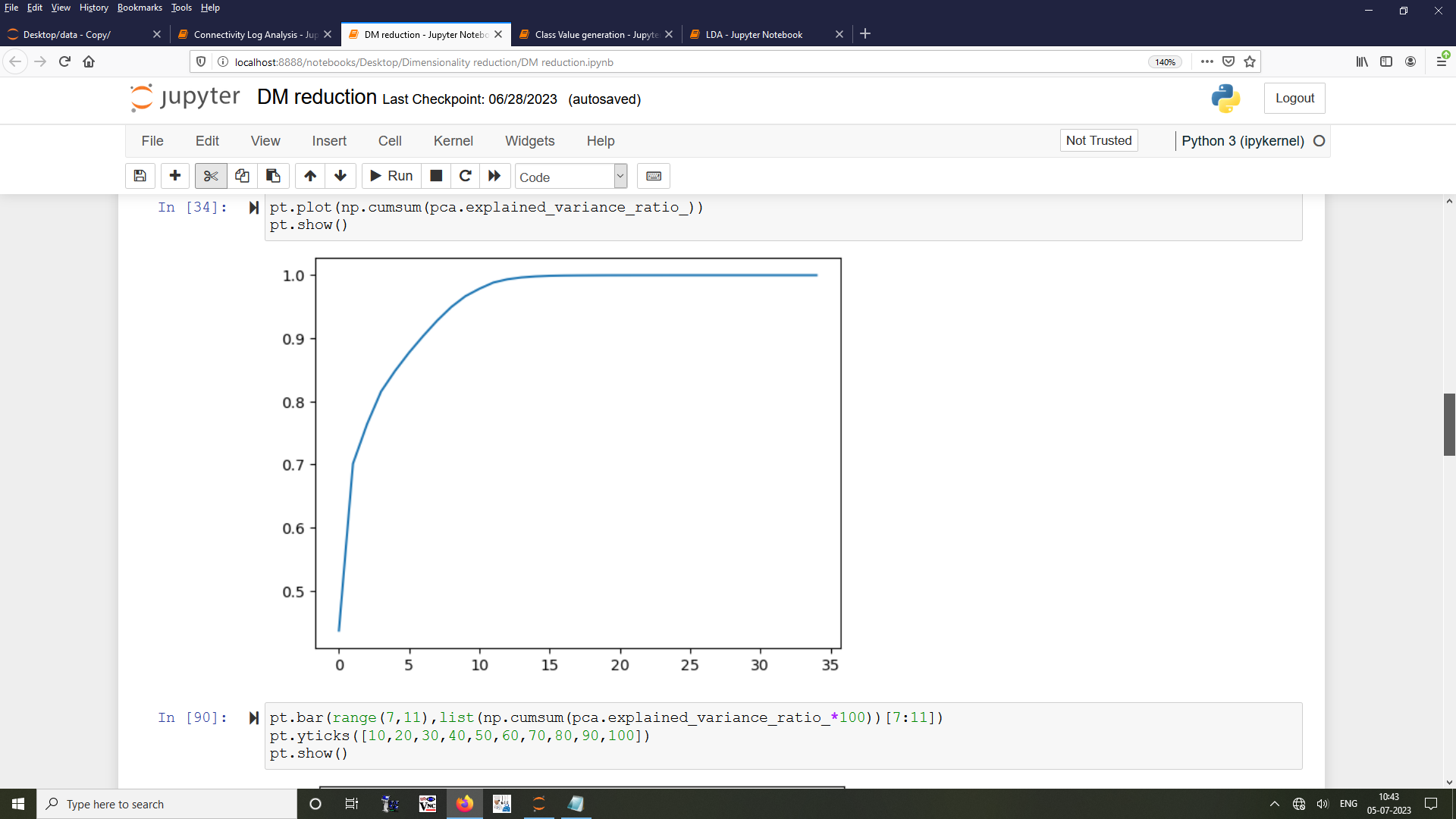
1. Load the engine data.



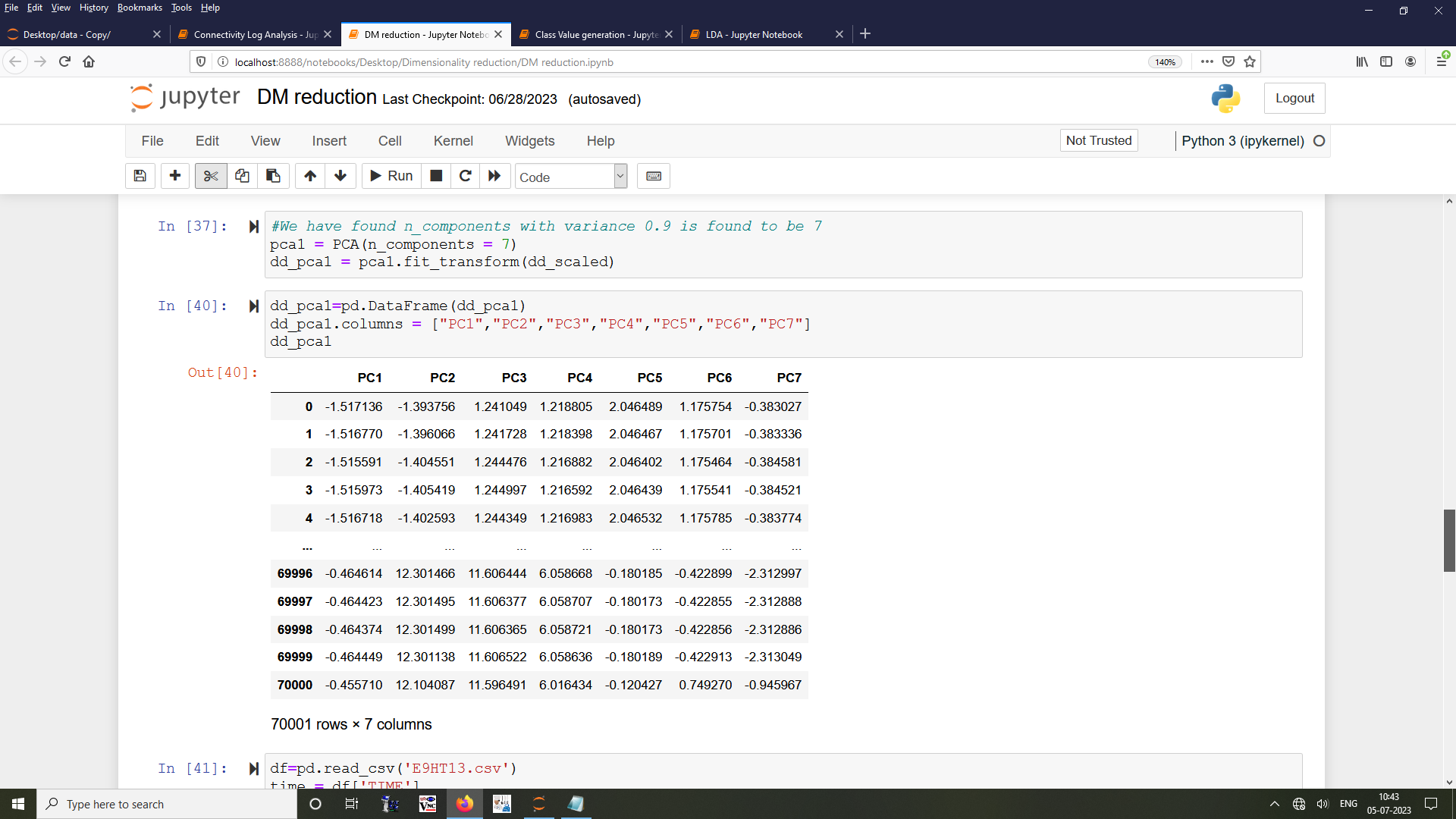
1. Apply standard scaler to the data (Z score).



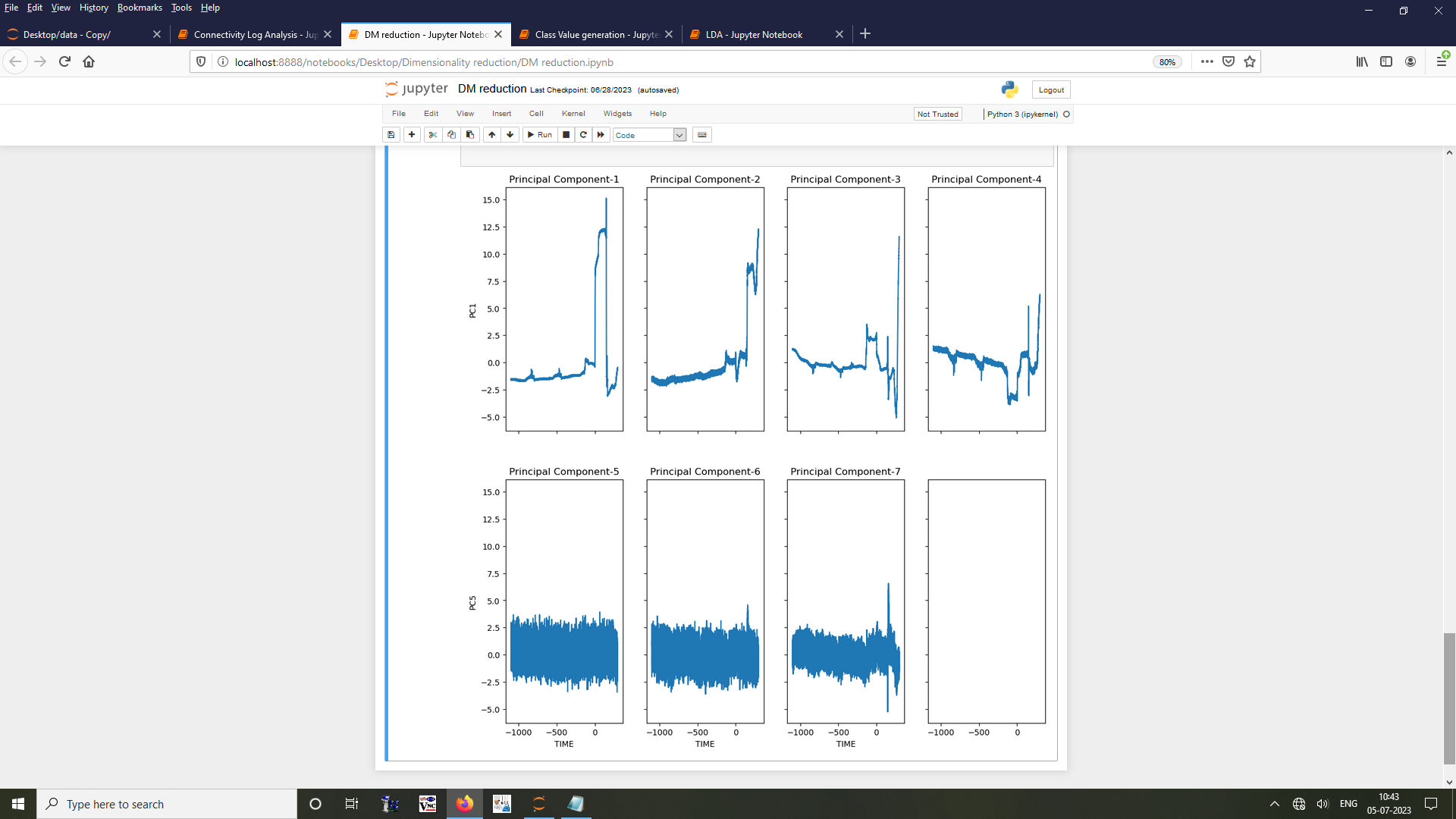
1. Find the principal components which explain 90% variance using line plot.



1. Perform PCA.



1. Show how the principal components vary when plotted with time.



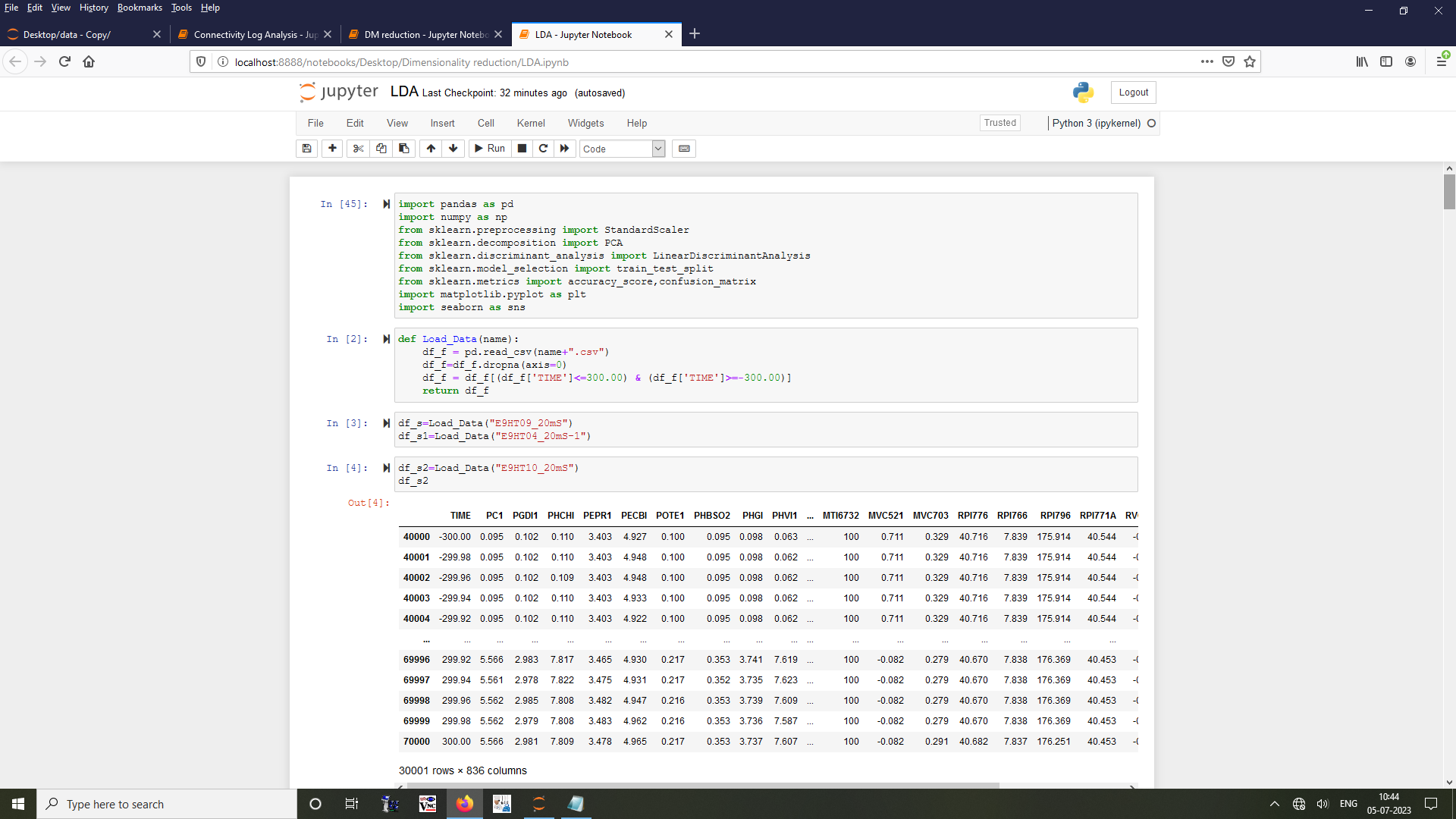
## 8.3 TASK 3 - Linear Discriminant Analysis:

**Aim:**

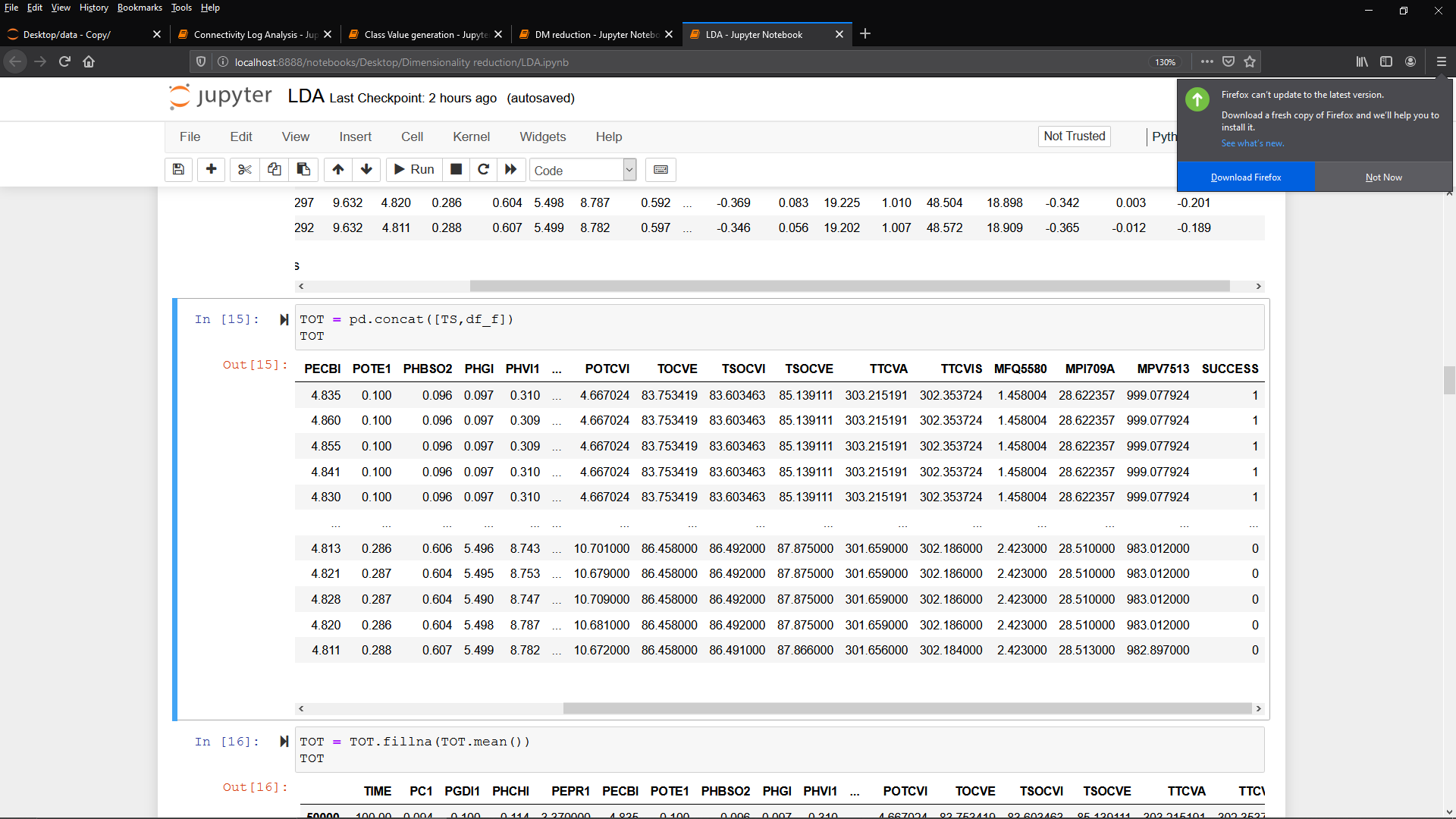
To perform dimensionality reduction using linear discriminant analysis (LDA) on given engine datasets and infer how the success-failure points vary from each other.

**Algorithm:**

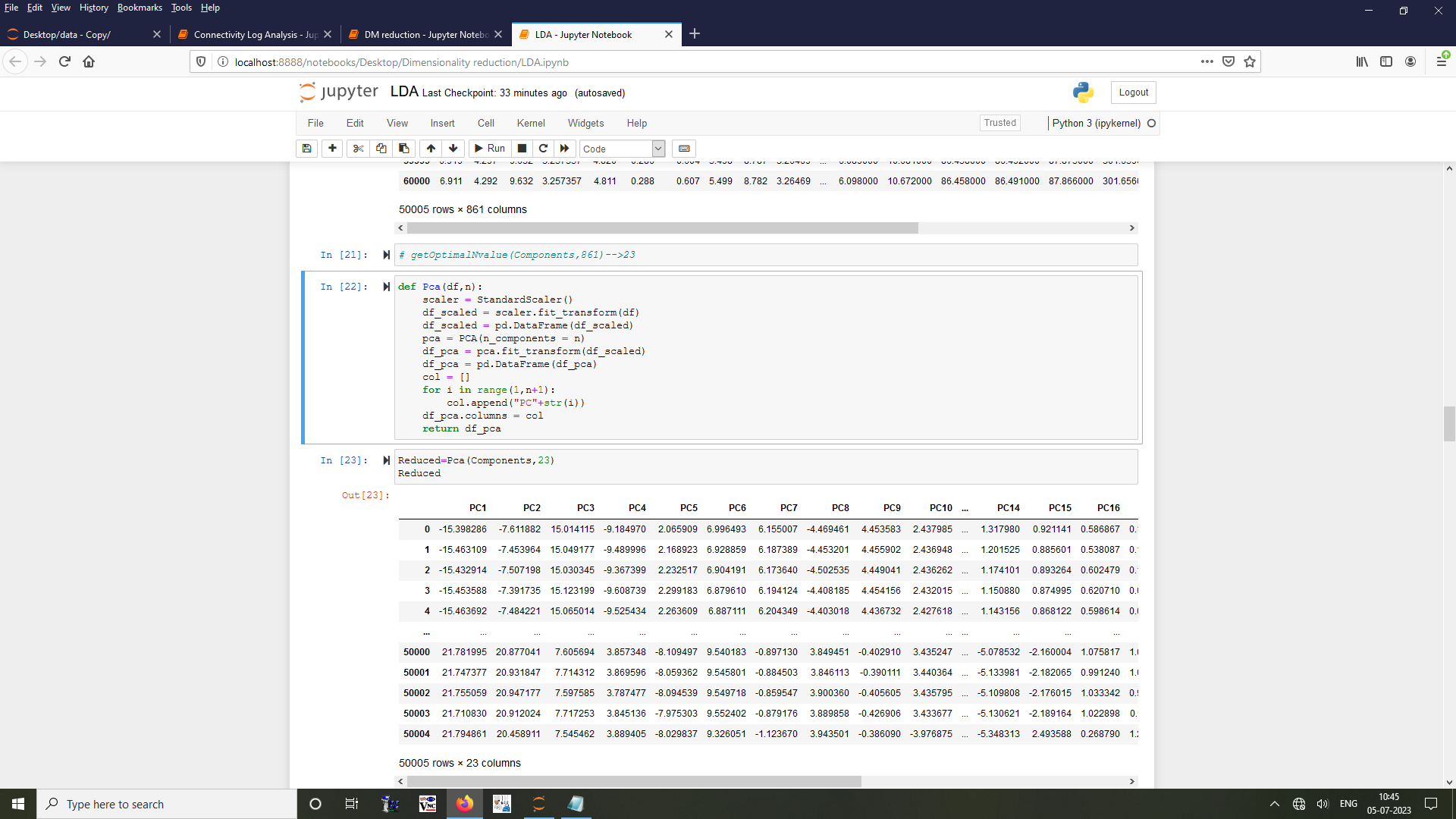
1. Load the datasets.



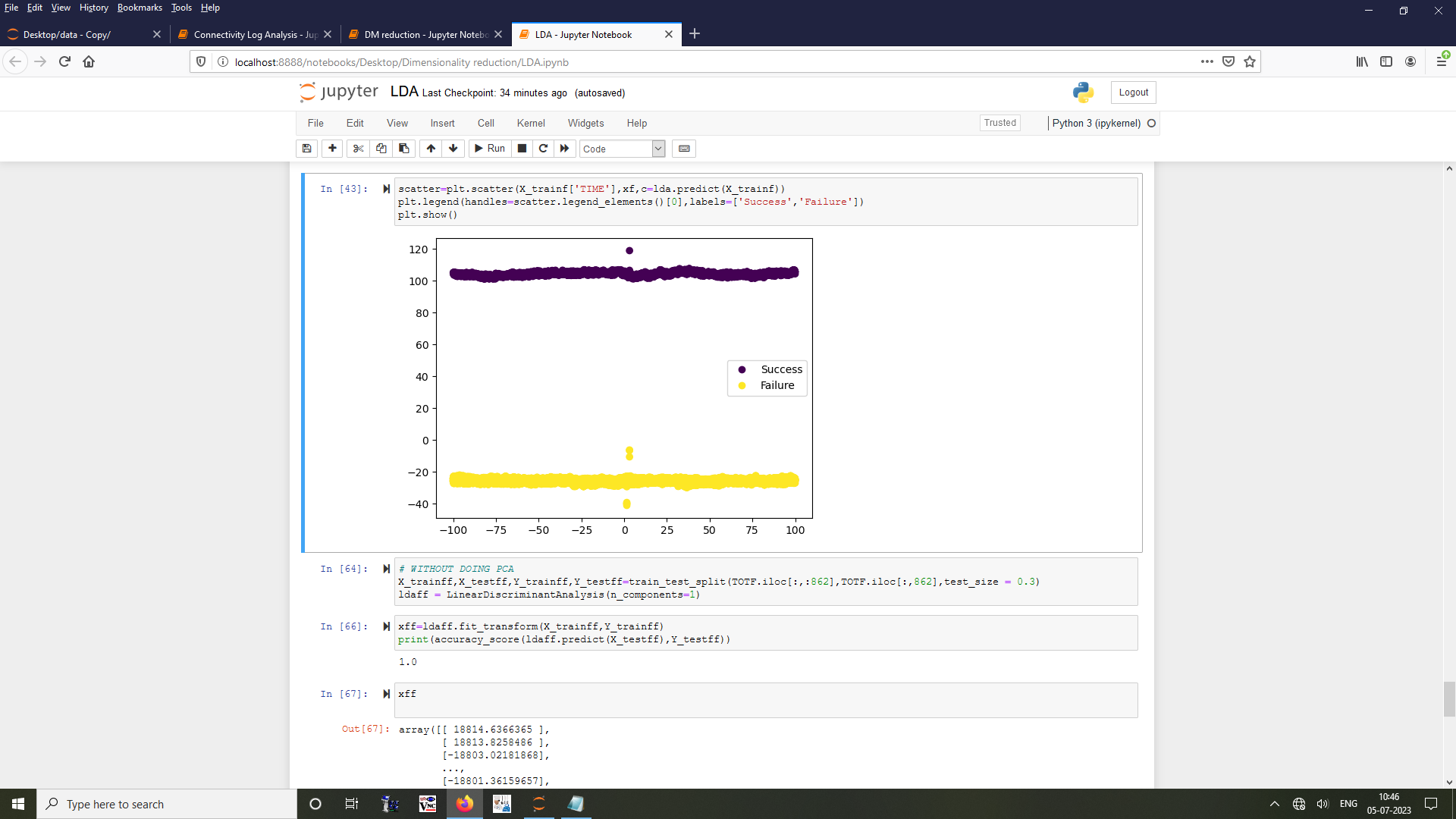
1. Merge them into single data after specifying the class value for them.



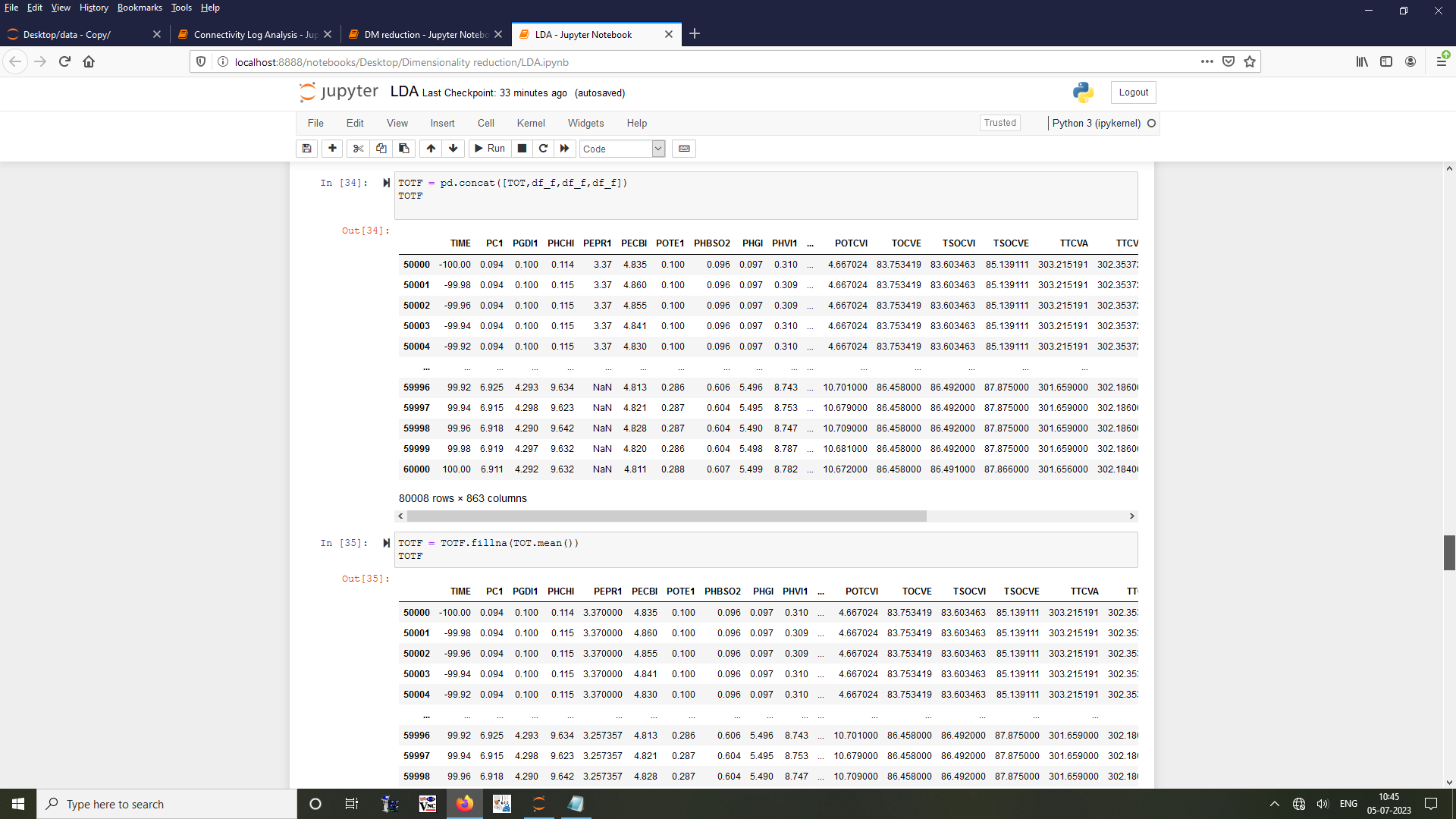
1. Perform PCA for the merged data excluding the Time and Class columns.

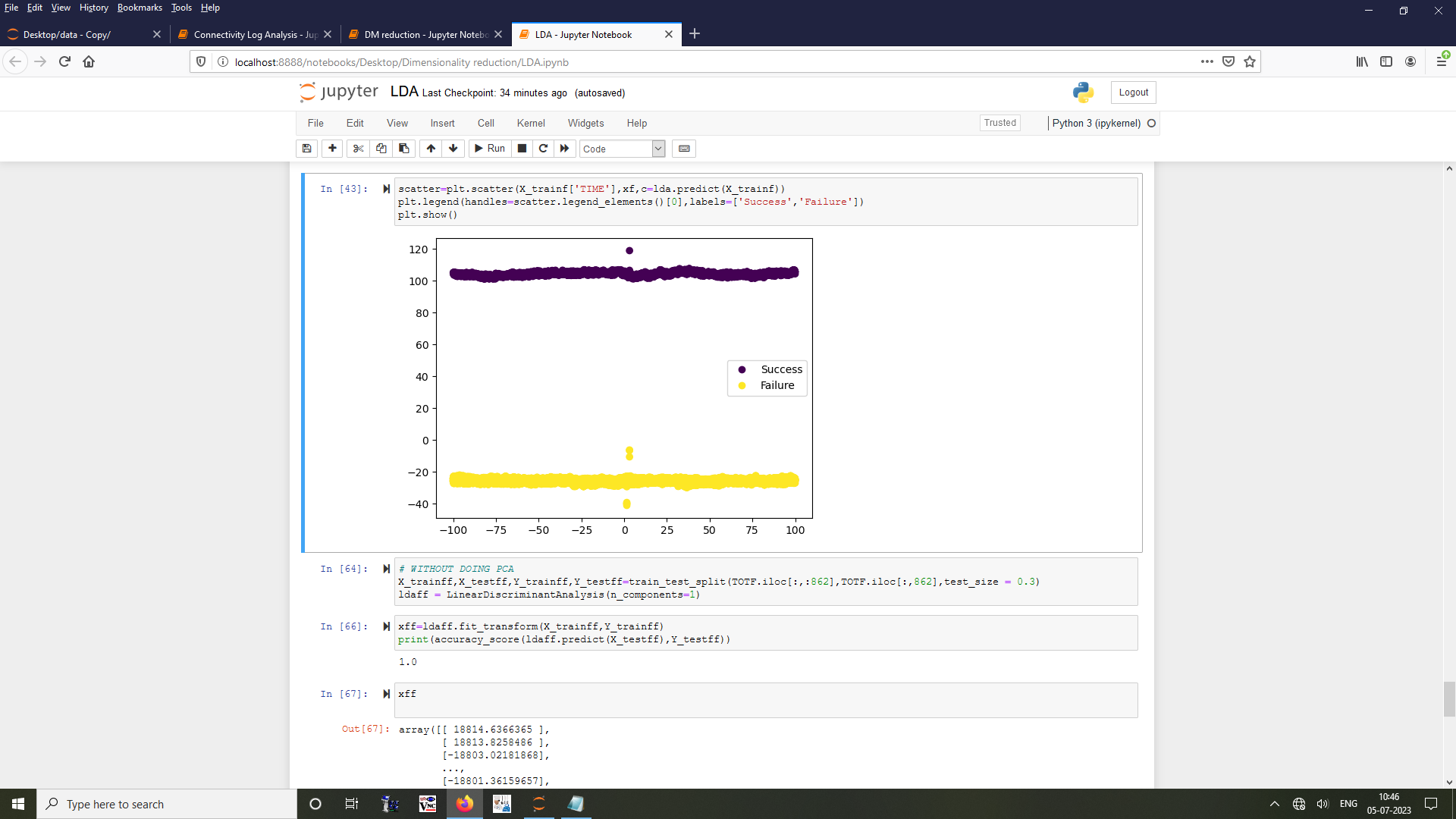


1. Perform LDA and plot the prediction on test data.

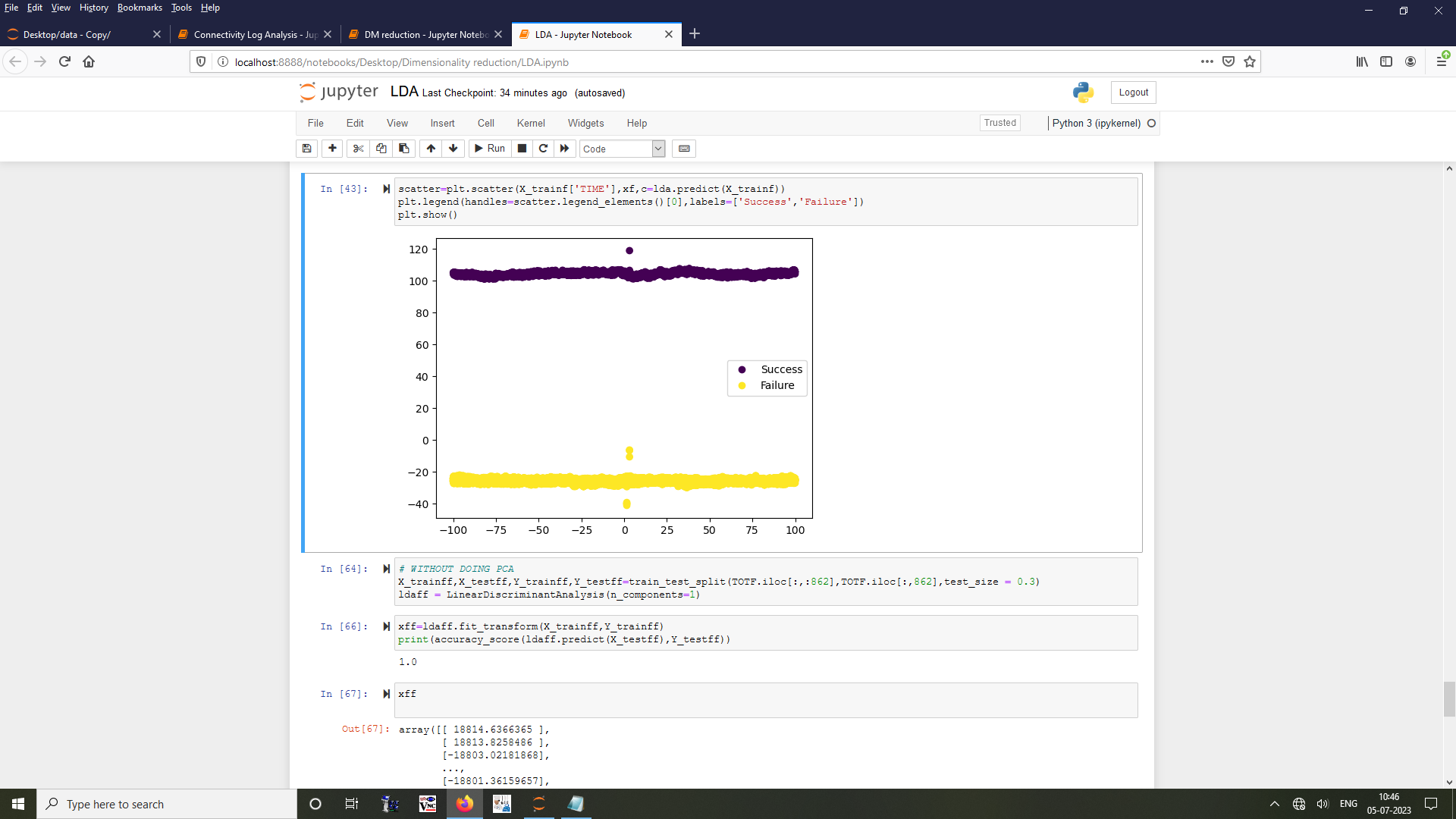


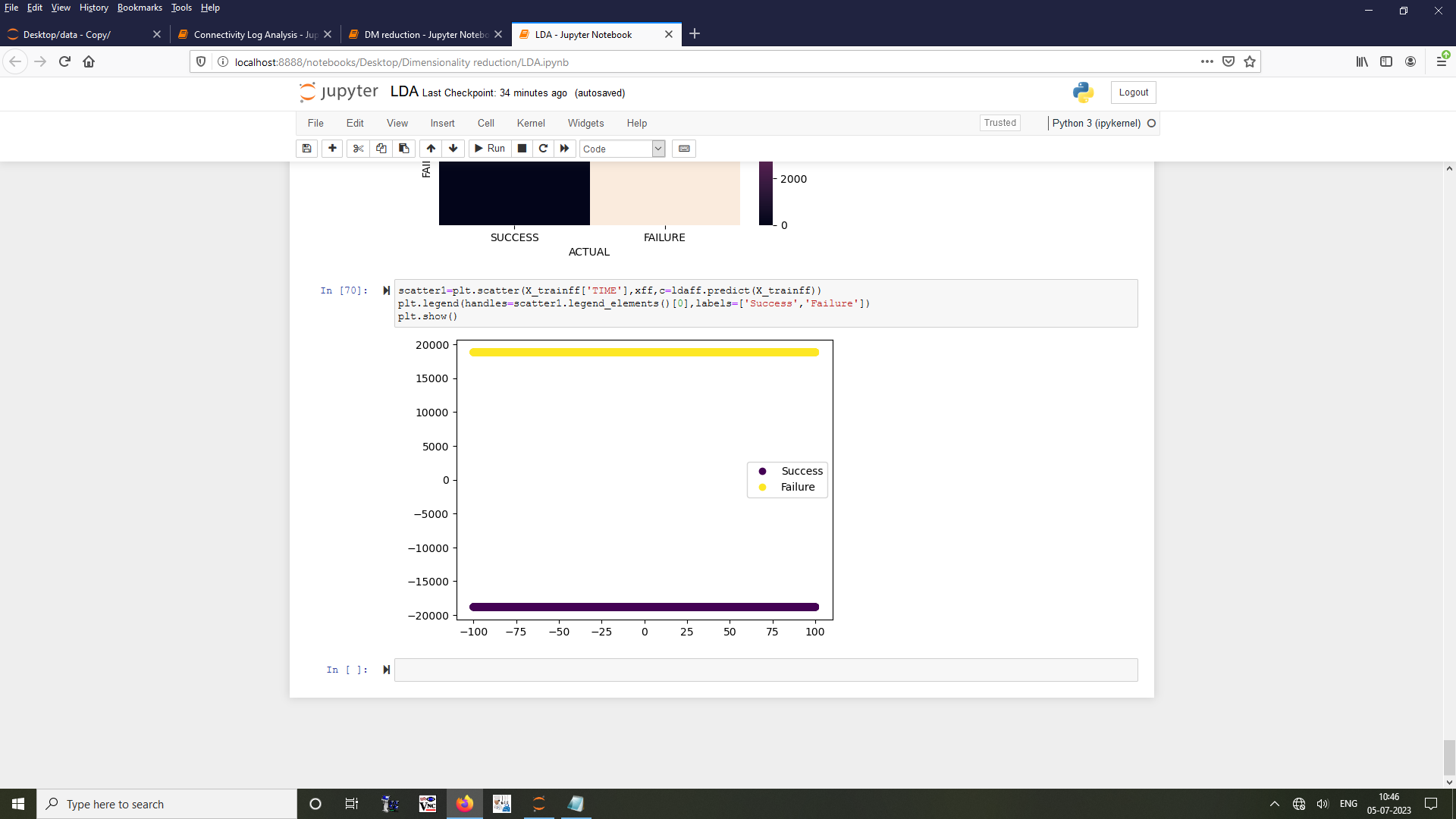
1. Increase the failure data thrice to avoid biasing and plot.





1. Perform LDA on actual data without performing PCA and plot.



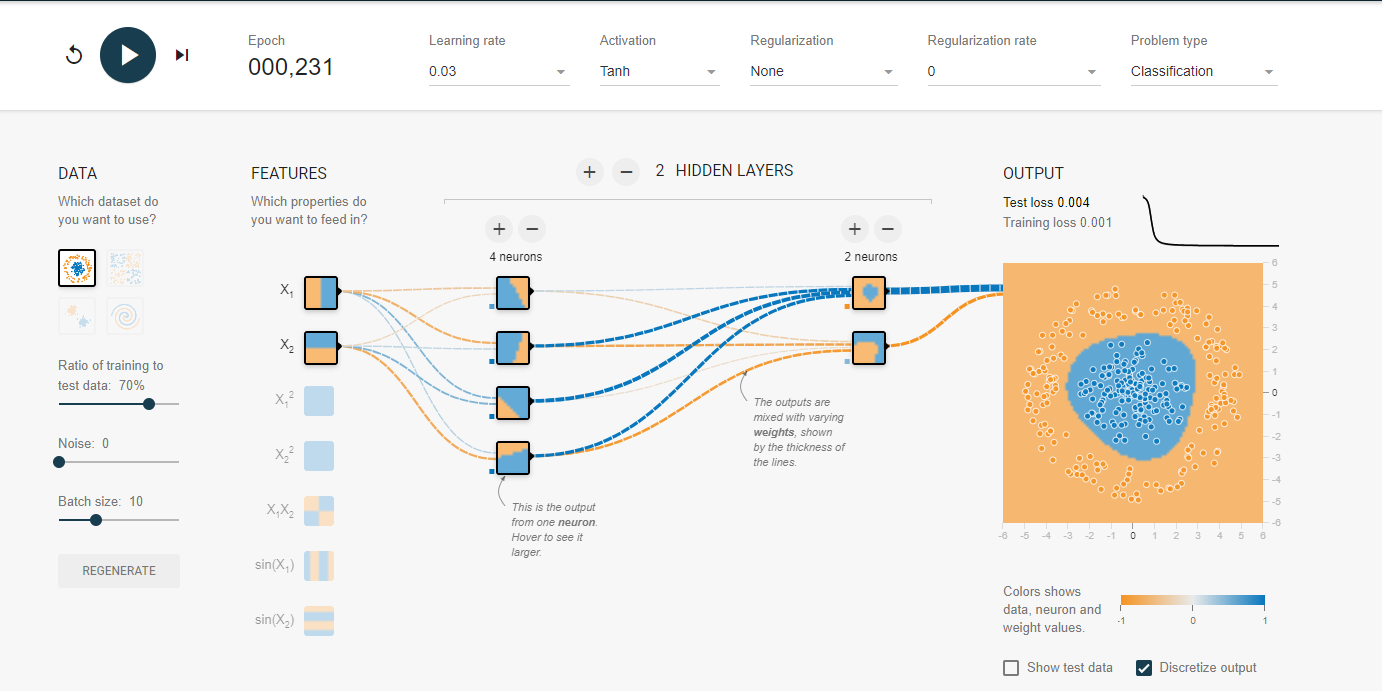


**Inference:**

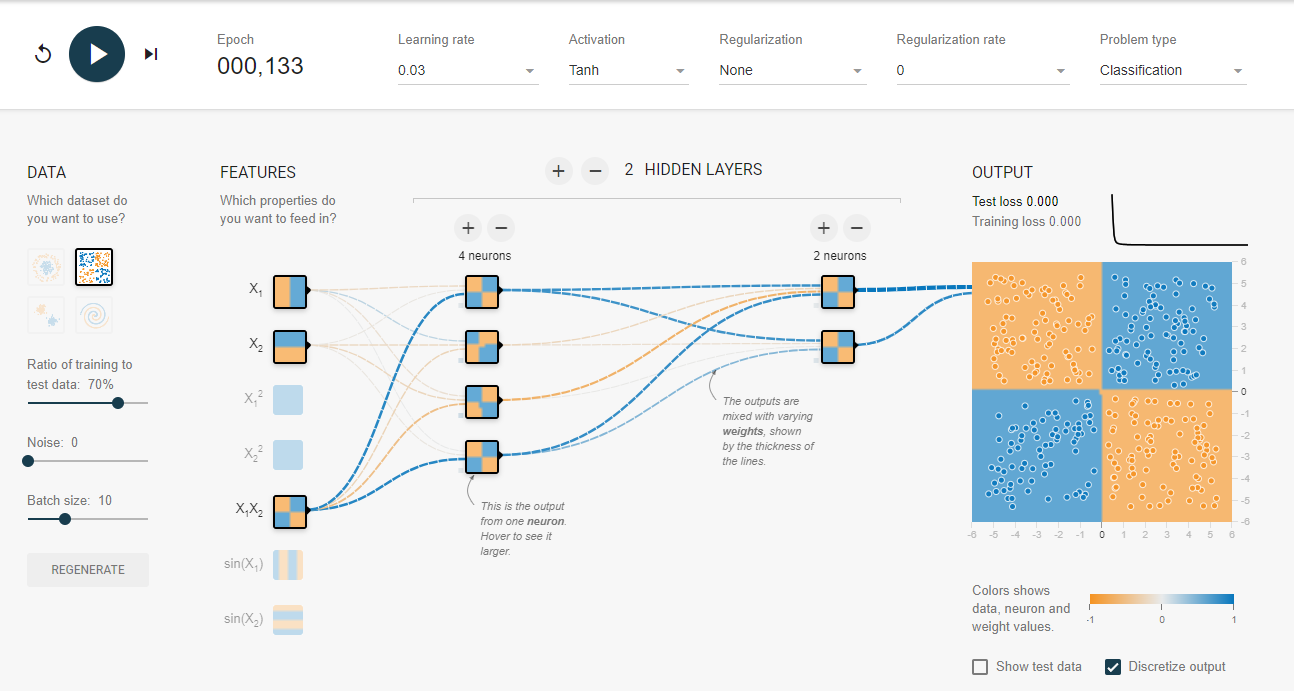
From performing LDA in this given engine dataset we could infer that in a time series data it is not effective to use LDA as it over fits the data to the model which cannot be used to predict data from other engine tests efficiently. So for this kind of data neural networks can be used to reduce the dimensions (Auto encoders and decoders).

## 8.4 TASK 4 – Neural Networks in Tensor-flow Playground:

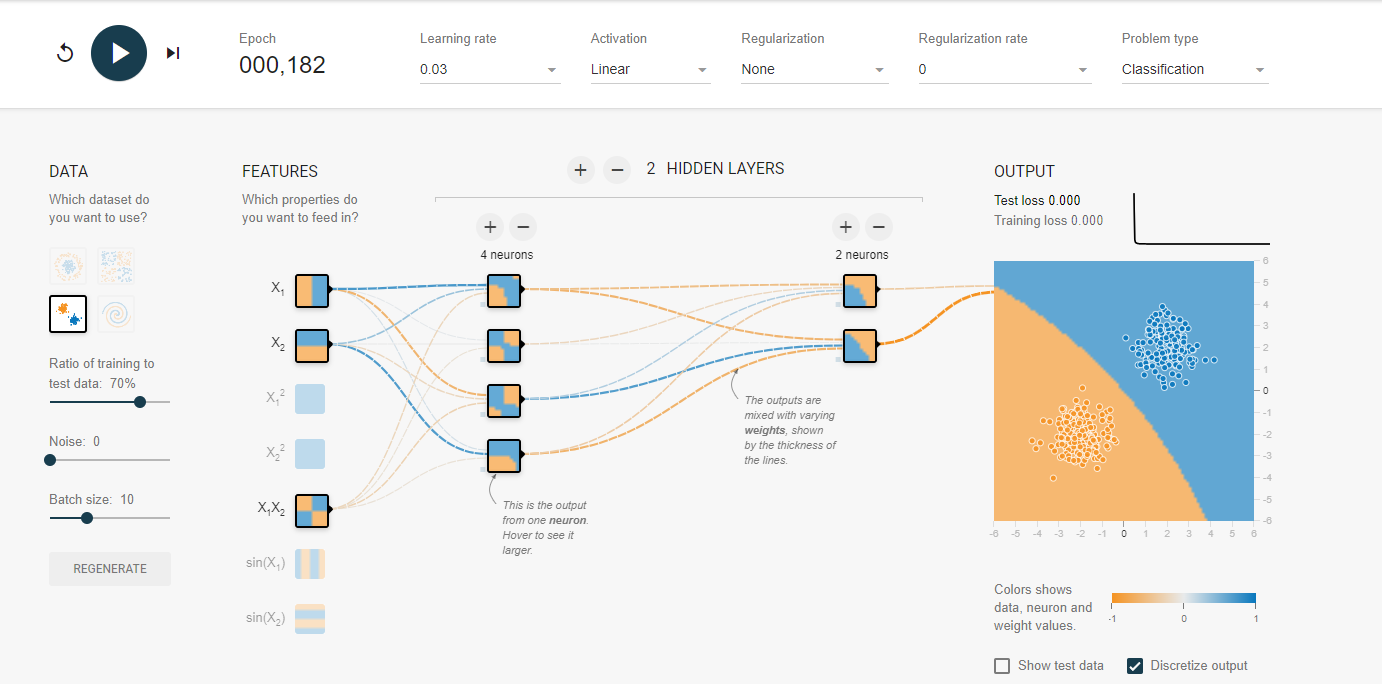
**Circle Dataset:**

****

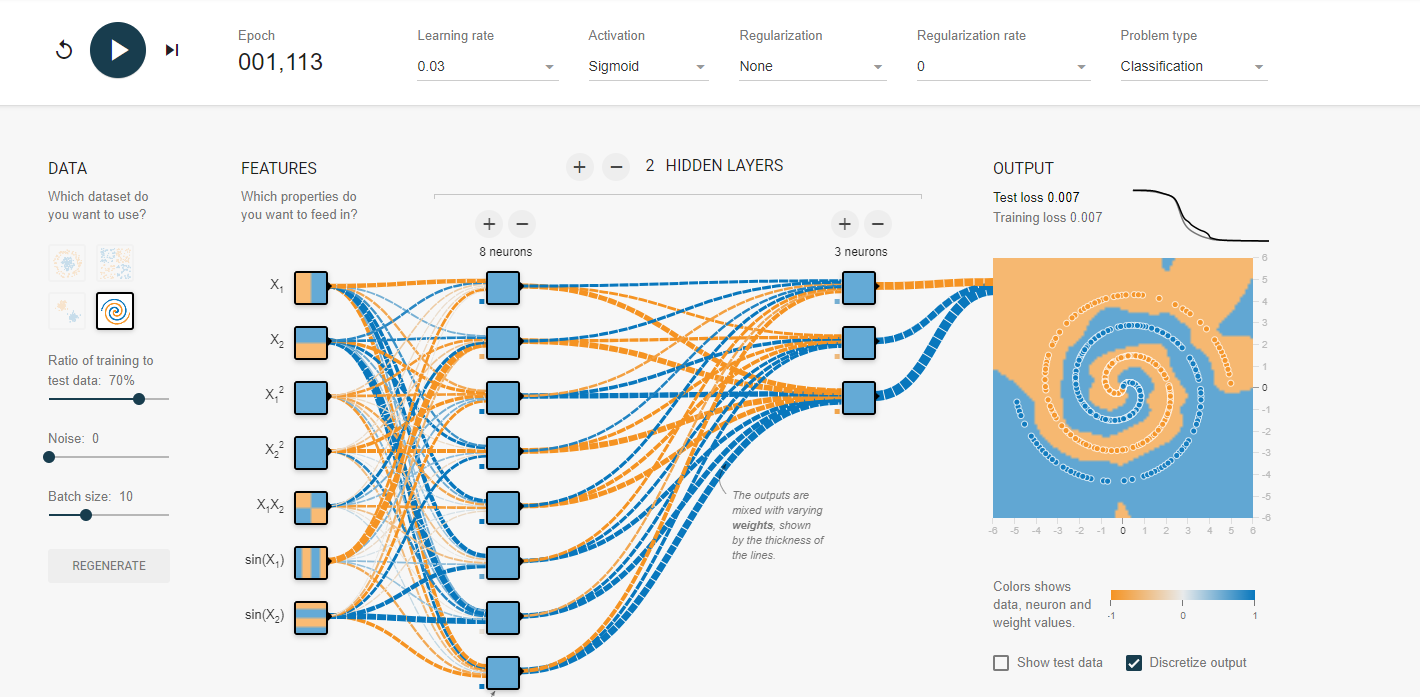
**Exclusive-OR Dataset:**

****

**Gaussian Dataset:**

****

**Spiral Dataset:**



# 

# 9. Conclusion:

It is a privilege to be able to attend this in-plant training in ISRO Propulsion Complex. We learnt about the need for data science to be applied in various fields. We learnt how different kinds of data are used to perform different tasks. The outcome of these tasks can be enhanced by using the data more effectively. We also learnt other systems in IPRC like Networking, Databases, Control centers, Servers, Telecom services and a basic outline of Test facilities. We also learnt different techniques to preprocess the data, reduce the dimensions as the data produced from the tests contain a lot of parameters and what techniques has to be applied for different data in order to get valuable insights. We hope that the knowledge we acquired will be used to further develop systems that use data efficiently.